

AUTOMATED DOCUMENT PRODUCTION FROM A SEARCH ENVIRONMENT

Technical Field of the Invention

The present invention relates generally to network-based browsing and searching applications and, in particular, to the preparation of printable documents that facilitate the rapid and discriminating review of search results obtained from such applications, and for providing advertising and like messages.

Background Art

In order to facilitate the accessing of information available through computer networks such as the Internet and the World Wide Web ("the Web"), network service providers typically allow users access to one or more search engines that are operable by the user to identify specific classes of information available on the network.

Fig. 1 depicts an example of the Web browser application and a typical presentation via a graphical user interface (GUI) of a search engine page 100 that may arise from a user searching of a particular keyword string of text 102, in this example STRING. STRING may be a single alphanumeric word, or a list of such words, perhaps linked by Boolean operators, and represents the search criteria used by the search engine. As illustrated in Fig. 1, search results 104 are typically ranked according to the quality of hit upon the searched string and typically lists a location 106 of the search result and the corresponding title 108. The actual presentation of the search results 104 is often user definable within a range of settings established by the search engine. Typically the location 106 of an individual result 114 is expressed as a Uniform Resource Location (URL). In some cases, the title 108 and the location 106 are combined as a Uniform Resource Indicator (URI). Often, the search engine application provides for some amount of text 110, typically a few lines or an abstract, relating to the particular location to be presented as part of the search result. In many search engines, this text typically represents the first few lines of text of the referenced location.

As a consequence, based on the presentation shown in Fig. 1, the user is then able to scroll through the search results 104 using a scroll bar 112 or by selecting different

display pages of the search result to identify those individual results desired to be reviewed. In those circumstances where the user preselects the text 110 to be displayed, often the search results 104 are provided arranged ten to a page. However, removal of the text 110 can provide the search results to be provided at a rate of approximately twenty
5 per page, where "page" in this fashion represents electronic page of information displayed on a display screen to the computer network user.

Many problems exist with the above described arrangement. Firstly, where only the URL 106 is presented in the result, often the user has no means of interpreting the search result other than by accessing the URL. Where the titles 108 are selected for
10 display, often the particular title 108 provides no information as to the specific content, or context of that content, to be found at the corresponding URL, or bears any relationship to the searched string 102. Further, where the text 110 is also provided, there is no guarantee that the searched string 102 will be presented in the text displayed on the search page 100. Further, in any configuration there is no guarantee that when the user accesses
15 the particular location 106, at which the searched string 102 is purported to be found, that the searched string will actually be found. As a consequence of the inadequacies of the information presented in the search page 100 shown in Fig. 1, users often spend excessive amounts of time accessing individual locations, reviewing the locations and, where appropriate, discarding the relevance of that location before referring to a further location.

20 Fig. 2A exemplifies a display of a Web page 200 accessed through a user selecting the first search result 114 shown in Fig. 1. As seen from Fig. 2A, the displayed Web page 200 includes a title banner 202, various images 204, 206, 208, 210, and text 212 incorporating the searched string 214. With such an example, accessing of the first search result 114 by the user provides an immediate result in response to the user's
25 searching of the search string.

However, in Fig. 2B, a Web title page 230 for another search result 116 shown in Fig. 1 is shown which also includes a title 216, a certain amount of text 218, an animated GIF image 220 together with a number of further URL links 222 within the same primary location depicted by the particular URL, in this case URL#m. Notably, the search string

is not seen in the displayed page 230 of Fig. 2B. In order for the user to find the search string, it is necessary for the user to scroll through the Web page 230 using the scroll bar 112. As seen from Fig. 2C, the search string 102 is located at 232 within a display screen 234 within the Web page 230 defined by the URL#m. The display screen 234 is
5 seen to have images 224, 226 and text 228.

Further, where the user reviews information at any one location, the only convenient way of forming a reasonable record of that review is to print a particular page of the reference location. Print facilities provided with browser applications and search engine pages are limited to one Web page at a time. This requires the user to access each
10 Web page and to then print that page where appropriate.

It will therefore be appreciated that traditional methods of viewing search results obtained over computer networks can be time consuming, and are not conducive to providing a convenient record of search results.

Summary of the Invention

15 It is an object of the present invention to substantially overcome, or at least ameliorate, one or more disadvantages of existing arrangements.

In accordance with one aspect of the present invention there is disclosed a method of presenting search results obtained from a search conducted over a computer network, said search including searching criteria and returning information including a
20 plurality of network locations, said method comprising the steps of:

- (a) extracting data from a first said first network location;
- (b) examining said data to identify therein said searching criteria to provide at least one specific location within said first network location of said searching criteria;
- (c) using said one specific location to identify from said extracted data
25 specific data including at least said searching criteria;
- (d) formatting said specific data into a printable document; and
- (e) repeating steps (a) to (d) for each remaining said network location in which step (d) appends said formatted data of said remaining network location to said printable document.

In accordance with another aspect of the present invention there is disclosed a method of formatting an electronic document intended for reproduction by printing, said method comprising the steps of:

- (a) sourcing main data from at least one location in a computer network, said data including a plurality of data types;
- (b) formatting said data into a common data type suitable for each of electronic display and printing;
- (c) arranging said formatted data as a printable document spanning at least one printable page;
- (d) identifying one or more locations where said at least one page is void of said formatted data; and
- (e) sourcing further data configured in said common type and sized to be positioned within said one or more locations; and
- (f) formatting said further data within said one or more locations in said printable document.

In accordance with another aspect of the present invention there is disclosed a method of formatting an electronic document intended for reproduction by printing, said method comprising the steps of:

- (a) obtaining from a searching process location information within a computer network of at least one search result returned by said searching process;
- (b) using said location information to fetch data from said computer network relating to each said search result, said data including said searching criteria; and
- (c) formatting the fetched data including said searching criteria into a printable electronic document.

Apparatus and computer program products for performing each of the methods are also disclosed.

Brief Description of the Drawings

A number of preferred embodiments of the present invention will now be described with reference to the drawings, in which:

Fig. 1 shows an exemplary view of a typical search engine page as seen using a Web browser application;

Figs. 2A to 2C depict example pages that may be found through traditional examination of the search engine page of Fig. 1;

5 Fig. 3 is a schematic exemplary representation of a search result page according to the preferred embodiment based upon the search result of Figs. 2A to 2C;

Fig. 4 is a flow chart depicting results collection and formatting according to the preferred embodiment;

10 Fig. 5 is a schematic illustration of a network traversal page according to another embodiment;

Fig. 6 is a flow diagram depicting generation of the arrangement shown in Fig. 5; and

Fig. 7 is a schematic block diagram representation of a computer system in which the preferred embodiments may be implemented.

15 **Detailed Description including Best Mode**

To assist users in being able to track and trace their traversal of computer networks such as the Web, Canon Information Systems Research Australia Pty Ltd has developed a "Hypertext Document Collating Tool" which is currently the subject of United States Patent Application No. 08/903,743 filed 31 July 1997. The Hypertext
20 Document Collating Tool operates in a background mode behind the browsing software application used to traverse the Web, so as to automatically and transparently create a printable document that includes the various Web sites and documents encountered by the user during the traversal of the Web. The Web sites and documents typically include numerous data types, components and configurations, such as simple ASCII text, JPEG
25 images, GIF and TIF static and animated graphics, and so on. Such varied source data is often termed "hypertext" and is formatted primarily for electronic representation via a display screen, but not necessarily for hard copy reproduction. Implementations of the Hyertext Document Collating Tool have been realised by the product marketed under the

trade mark WebRecord™ by Canon Kabushiki Kaisha and Canon Information Systems Research Australia Pty Ltd.

The embodiments described in the present specification are preferably implemented as additional features to the Hypertext Document Collating Tool. The present invention however is not limited to user with the Hypertext Document Collating Tool, WebRecord™, or similar products, but has wider application and may for example be implemented in generic browsing software or searching arrangements, as will be appreciated by those skilled in the art having read and understood this specification.

The preferred embodiment provides an arrangement whereby the user, having conducted a search using traditional browsing software or search engines, such as described with reference to Fig. 1, is not required to individually examine each of the locations identified in the search result. According to the preferred embodiment, a further application, herein referred to as a formatted document generator application, operating in a manner akin to the aforementioned Hypertext Document Collating Tool, receives input information from the search engine result in the form of the searched keyword string and the individual URLs of each search result, and thereafter automatically generates a printable document formed of extracts from each search result, each extract incorporating information representing the context of the individual search result together with further information indicating the specific content incorporating the searched string. The input information may be provided directly from the search engine page 100, or alternatively via a file, for example pointed to by the search engine page 100, that contains the input information in a format interpretable by the formatted document generator application.

Fig. 3 shows a window 300 forming part of a graphical user interface presenting an electronic document 302 to the user. The electronic document 302 is presented formatted for printing, and includes two printable pages 304 and 306 scrollable using a scroll bar 308 and thus represents a "print preview" type display as would be understood by those skilled in the art. The printable page 304 of Fig. 3 is formatted into two columns 310 and 312 thus proportionally compressing the size of individual components incorporated on the page and thereby increasing the amount of information that may be

reproduced on a single printable page. Each of the columns 310 and 312 is divided into sections 314, each section 314 relating to a specific result 104 indicated from the search page of Fig. 1. The sections 314 are separated by first graphical dividers 316 inserted by the formatted document generator application. Each section 314 is divided into two portions 320 and 322 by means of a second graphical divider 318 also formed by a graphical object inserted by the formatted document generator application.

The first one 320 of the portions incorporates an extract of the top of the Web page associated with the URL of the corresponding search result. As seen, the top of the first column 310 of Fig. 3 corresponds with and replicates the top of the Web page shown in the display of Fig. 2A. This information is indicative of the context of the particular Web page provided at URL#1 and thus affords the user a context for the information contained in the search string 102. The portion 320 below the graphical divider 318 incorporates that portion of the Web page which includes the searched string 102 and thus provides the specific content relating to the string as it is found in that Web page. With respect to URL#1, this will be seen to be consistent with that substantially displayed in the arrangement of Fig. 2A.

This separation and presentation of information is more generically depicted in Fig. 3 with respect to URL#2 with respect to the top of the Web page context information being separated from the content information including the searched string.

The example shown in Fig. 3 with reference to URL#m illustrates the specific power of the preferred embodiment in avoiding a need of the user to review and/or print the entire contents within any one search result URL location. As seen, the section relating to URL#m includes a contextual title and page corresponding to that shown in Fig. 2B, followed by a sectional divider- 318 and a section corresponding to the information contained in Fig. 2C. Significantly, use of the formatted document generator application permits passing over, or skipping, those portions of the referenced location (in this case URL#m) that do not contain the searched string. In some situations, this can obviate the necessity to scroll through or print a single Web page that may occupy many display screens of information, which may also occupy many individual sheets of paper in

a hard copy reproduction. The preferred embodiment acts to condense the particular information found at the location identified by the search result into a convenient, interpretable and manageable form.

As seen, the printable document 302 depicted in Fig. 3 incorporates information
5 from a number of Web locations identified by the search result, the information being presented in a contextual form and also in a form in which the particular content relating to the search string is identified.

As also seen in Fig. 3, each of the portions 320 includes a reproduction 324 of the actual URL to which that portion relates thereby maintaining a record of the location
10 of the referenced information. The portions 322 may also be configured to identify a similar location 326, but in this case, modified by the relative location within the referenced URL as to where the searched string is found. For example, for URL#1, the searched string is found on "page 1", representing the electronic screen page number corresponding to that seen in Fig. 2A. For URL#m, the page number is "page n",
15 representing the number of electronic screens required to be conventionally scrolled by the user within URL#m in order to locate the searched string.

A permanent record may be obtained either by saving the electronically displayed printable document 302 to memory, for example by actuating a SAVE icon 324, or by printing the electronically displayed printable document, for example by
20 actuating a PRINT icon 326.

According to the preferred embodiment, the printable document of Fig. 3 can be generated once the search engine results of Fig. 1 are returned to the user via the browser application. Generation of the printable page of Fig. 3 occurs as a result of the hypertext document collating and formatting, and without additional burden upon the Web browser
25 application or search engine, by directly and separately accessing the individual URLs returned by the search engine result and searching within each URL for the relevant searched string information. This method is depicted in the flow chart of Fig. 4.

Fig. 4 depicts the operation 400 of a Web browser 402 and associated search engine 404 and their relationship to the operation of the formatted document generator

application 430 according to the preferred embodiment. As seen, the browser 402 incorporates the search engine 404 which acts upon a search string 406 entered by the user. The search string 406 may be one or more words of text separated or associated using Boolean operators. The search engine 406 returns a search result 408, typically
5 being an individual page displayed by the browser 402 and which traditionally replicates the searched string together with an indication of the total number of results 410 or hits identified by the search. The user is then able to select groups 414 of results typically based upon a ranked number which can be displayed on a page at any one time. As seen, results 412 incorporating Results # 1-m are displayed.

10 At this stage, the user may invoke the operation of the formatted document generator application 430 to create a printable document based upon the search result 412 indicated in the search page. If this is not desired, operation of the application 430 is not performed and the user is free to continue utilising the search engine 404 or browser 402 in a traditional fashion, for example by effecting a further search or directly examining
15 any one of the search results 412. Hypertext document collating and formatting may be performed if such is desired.

Alternatively, and according to the preferred embodiment, where the user elects for the creation of a printable search result in step 432, step 434 checks that the printable document is to be formed for the displayed results, in this case, Results # 1-m. Where
20 appropriate, the user may select, via an interconnection 422 to the search result 408, for another group 414 of the search results 412 to be selected. Once the group of results is settled in step 434, step 436 copies search information data including the searched string 406 via an interconnection 418, and the URL's corresponding to the selected group of results 412 via an interconnection 420.

25 The interconnections 416, 418, 420 and 422 shown in Fig. 4 will be appreciated as being illustrative of interactions between the applications 402 and 430, and those skilled in the art will appreciate that such may be implemented in various forms or procedures, not necessarily dependent on the individual "interconnections" as shown, any one of which may be optionally implemented. For example, interconnections 418 and 420 may

be unitarily formed as a single request for search information, such as when the search information is derived from a file, the file for example being formed by batch processing of search engine calls.

In step 438 which follows, the generator application 430, independently of the
5 browser 402 or search engine 406, fetches the data from the Web at the various URL's
given by the individual results 412. Step 440 then commences a processing loop on the
fetched/downloaded data, one result at a time. At this stage, a determination is made in
step 440 as to whether or not there are any unprocessed results and, where there exist no
are unprocessed results, step 442 follows. In step 442, examination is made of the fetched
10 data of the particular result 412 to identify if the searched string 406 is actually found
therein. In those cases where the string being searched is not found within the Web site, a
situation which occurs all too frequently and much to the displeasure of Internet users,
step 442 returns control to step 440 so as to process the next result. Where the searched
string is found in step 442, step 444 follows which records the specific location of the
15 search string 406 within the particular Web site.

Processing of the individual result location is then performed to format the
search result. Next, in step 446, the top of the Web page for the result being processed is
extracted from the fetched URL data. In step 448 this is converted from a hypertext
format to a common format suitable for both electronic display and hard copy printing,
20 and incorporated into a printable document. Step 450 which follows inserts a result
divider, corresponding to the graphic 318 of Fig. 3, into the formatted document.
Step 452 then uses the recorded location from step 444 to extract a particular section of
the Web page that incorporates the searched string. In step 454 which follows, the
extracted selection incorporating the searched string is formatted into the common
25 displayable and printable form and incorporated into the printable document. This is
followed by step 456 which inserts the location break, corresponding to the separator 316
of Fig. 3, into the printable document. Control then returns to step 440 for the processing
of the next result. Where there are no further results to be processed, step 440 transfers
control to step 458 which enables the user to view the printable document via a display

screen in the fashion shown in Fig. 3, from which search results may be saved and/or printed by actuating an appropriate icon.

In an alternative implementation, steps 438, 442 and 444 may be combined into a single process where, as the data is fetched and downloaded from the Web, the data is simultaneously checked to identify the searched string, and where the searched string is identified, that specific location within the data is recorded. Where the searched string is not located in the URL accessed data, that data may be discarded, without further processing, so that data from the next URL in the search results list can be fetched.

According to the preferred embodiment, the formatted document generator application extracts only predetermined portions of each search result URL thus ensuring that the printable search result document 302 of Fig. 3 does not contain extraneous matter not necessarily related to the search. This avoids consuming inordinate computing resources for generation and reproduction of the printable document.

Once reproduced, by display or printing, the user may closely review the search result presented by the printable document 302 and thereafter access particular URLs that may be desired for closer examination and review. A traditionally formatted printable document may then be created for any closely examined URL, according to the principles of the Hyertext Document Collating Tool.

Where desired, the formatting performed in step 454 of the extracted section incorporating the searched string may be made in such a manner so as to highlight the particular search string as it appears with in the formatted document. In this fashion, any person reviewing the printable document of Fig. 3 either in its electronic or hard copy printed form, has their attention directed to the particular search string as it is reproduced.

It will be further appreciated that the purpose of the search is to identify the searched string, and thus the display of the extracted top of page for the result is not essential for the performing of the present invention since this is not necessarily important to the particular search result. However, the present inventors consider that the incorporation that such of page information is relevant so that the actual searched string as found is placed in some user interpretable context. For example, a search of the term

"automobiles" may return results relating to the manufacturers of automobiles configured for use on the road. However, the same search may return a result for a manufacturer of toy or model automobiles suitable for the playing of children's games and the like. In many instances, the names of such automobiles and trade marks associated with such automobiles may be the same, irrespective of whether they are real automobiles or toy automobiles. The incorporation of the contextual top of page information will generally assist the user in distinguishing between toy motor vehicles and real motor vehicles.

The printable document generated according to the preferred embodiment may be arranged to extend over one or more printable pages which can be viewed via the display shown in Fig. 3 using the vertical scroll bar 308. Since the GUI display of Fig. 3 is a "print preview" type display, it has a fixed width and a horizontal scroll bar is not necessary. Further to the user's selectability of the display of the top part of each Web page, the preferred embodiment may be configured for the display of more than one section which incorporates the searched string, each such section for example being separated by replication of the graphic divider 318. For example, preferences may be established for reproducing a predetermined number of hits upon the search string and/or all of the hits on the search string.

Further, the amount of information reproduced relevant to any one hit on the searched string, and as a consequence of step 452, may be varied according to user preferences. By default, such may include a single displayable screen taken from a Web page incorporating the searched string. Alternatively, the entirety of the Web page at a particular location may be referenced. In a further alternative, where the text string forms part of a paragraph of text, only that paragraph or section of text need not be extracted for reproduction. In a further user selectable optimisation, because many images found on the Web are not directly reproducible in print, examples of which being animated GIF images and the like, the user may select disablement of such images in the printable document thereby permitting the formatted document generator application to reconfigure the printable document compared to the actual Web produced location so as to optimise the amount of text to be reproduced. Other image types, such as JPEG and static GIF

images may be disabled from printing as desired. This may be important, particularly when handling "home pages", which represent a root directory URL. In such pages, and many others, user selectable icons and the like often consume much of the displayable page but often provide no substantive information, particularly in satisfying a search query. Appropriate configuration of the formatted document generator, can permit such icons and like objects to be excluded from the printable document, thus affording greater levels of compaction of information relevant to the searched string in the printable document.

As a consequence, the printable document 302 of Figs. 3 and method of Fig. 4, permit network obtained search results to be automatically formatted and collated into a single, user interpretable, document that provides for compaction of more than one, and typically numerous, summaries of each search result on any single printable page. The embodiment also provides a permanent record of the search and search results in a convenient and compact form. Significantly, the information is presented to the user in the same manner that such would be seen had the user actually accessed the referenced URL in the traditional fashion, thus ensuring maintenance of the context of the referenced URL and the searched string to the user. This is to be contrasted with prior art search result presentations such as Fig. 1, which are not configured to reproduce graphics, images and other indicia that provide user interpretable context to the search result. Prior art search result presentations typically carry little or no information of the searched string.

A further embodiment is illustrated in Fig. 5 where a further printable page is illustrated as part of a GUI incorporating sections configured according to the previous embodiment.

As will be appreciated by those skilled in the art, automated collating and formatting of the printable documents of Fig. 3 can result in the content of any printable column or page extending to a further column or page thus causing the printable document to be somewhat non-contiguous. Whilst such may be often tolerated, where the document is segmented in the fashion shown in Fig. 3, it may be desirable for each

individual segment to be formed in a single column upon a single page. Such can result in the creation of vacant spaces at the end of columns into which no specific information may be placed due to size limitations. As seen from Fig. 3, such vacant space appears at the bottom of the column 312.

5 According to the further embodiment of Fig. 5, such vacant spaces may be occupied through the insertion of a printable message such as an advertisement. The message or advertisement may be inserted by a server associated with the generation of the printable document and/or the information contained therein. In a preferred implementation, as seen in Fig. 5 where a printable document 500 generated is formed
10 from search results according to the embodiment of Fig. 3, the printable document 500 includes a page break 502 separating a first printable page 504 from a second printable page 506, and a first column 508 into which are neatly placed two search result sections 510 and 512. However, in this example, the search result returns only one further result 514 which does not occupy the entirety of a second column 516 of the
15 page 504, leaving spaces 518 and 520 respectively at the top and bottom of the column 516. According to the present embodiment, advertisements are sourced based upon the subject matter being searched, in a manner consistent with traditional electronic advertisements that can occur during Web browsing sessions, the sourced advertisements being formatted into a printable form and inserted into the spaces 518 and 520. However,
20 unlike such traditional Internet advertisements, which are configured for transient display upon an electronic display (eg. video screen), advertisements of the present embodiment are intended ultimately for reproduction by means of printing and, as a consequence, are specifically configured for such display and will therefore, for example, be absent animated graphical objects or other moving components. The data format of such
25 message is therefore preferably common to that used in sections 510 and 512.

Using the foregoing example, where the user searches the string "automobile", an advertisement 522 may be placed in the space 518 advertising "AUSSIE Motor Vehicles" in the manner illustrated in Fig. 5. Another advertisement 524, for example

relating to automobile parts such as those manufactured by "JIM's Mag Wheels", may be inserted into the otherwise vacant space 520..

According to the present embodiment, one or more of the printable messages or advertisements may be returned to the formatted document generator in a number of ways. Firstly, this may be by way of the particular search engine being used to conduct the search on the string which traditionally returns electronic messages for display of the video screen of the user. Any such returned advertisement may then be interpreted by the formatted document generator which then converts into a printable form suitable for formatting and placement into the printable document of Fig. 5. This method, and another, may be described with reference to Fig. 6.

Fig. 6 shows a system 600 in which a user 602 operates a Web browser application 604 together with a formatted document generator application 606. The user 602 accesses a search engine server 610 via a network 612 in order to identify information available via the network 612. The search engine 610 is associated with an advertisement server 614 which includes a keyword algorithm interpreter 616 for examining the user's search string to thereby return an appropriate electronically displayable advertisement 622 via the search engine server 610 to the user 602. The formatted document generator 606 then interprets the advertisement and formats the advertisement into a form suitable for electronic printing by depositing the same into an electronic printable document 608, which, as will be observed, corresponds to the representation of Fig. 5.

In an alternative, the formatted document generator 606, having identified the need for the placement of an advertisement to fill the vacant column space, is configured to communicate via the network 612, without interaction with the browser 604 or search engine server 610, with a dedicated print advertisement server 618 for the provision of a printable advertisement to the formatted document generator 606. The print advertisement server 618 is configured in a fashion to examine the network 612, for example including the advertisement server 614 and/or itself, to identify one or more printable advertisements 620 to be automatically returned, preferably via the server 618,

to the formatted document generator 606. In this fashion, rather than relying upon the search engine server 610 for the presentation of advertisements, the formatted document generator 606 may call the dedicated server 618 to provide advertisements specifically configured for reproduction by both display and printing. In this fashion, the formatted document generator 606 either directly, or via the server 618, may control those advertisements that may be placed into the otherwise blank spaces 518 or 520 of the document 500. Such an arrangement, where calls are made via the server 618, permits monitoring of advertisements returned for printing in the document 608.

A specific advantage of incorporating the printable advertisements in the printable document is that the printable document can become a permanent record of the user's search of the Web that may be required for later review. As a consequence, the advertisement that is printed with the search result also becomes a permanent record of the advertisement and thus can be interpreted as being substantially more valuable in an advertising sense than a transient advertisement, such as those traditionally displayed on electronic display apparatus via the search engine server 610 as will be known to those skilled in the art. As a consequence, the individual printable advertisements can be provided at a premium cost compared to those transient advertisements. Further, where the formatted document generator 606 interacts with the server 618 for accessing the printable advertisements 620, the dedicated server 618 may be configured for direct management of costs associated with advertising charges associated with providing the printable advertisement 620 to the user 602 for incorporation into the printable document 608.

Formatted document generation described with reference to Figs. 3 to 6 is preferably practiced using a conventional general-purpose computer system 700, such as that shown in Fig. 7 wherein the processes of Figs. 4 and 6 and the Hypertext Document Collating Tool may be implemented as software, such as an application program executing within the computer system 700. In particular, the steps of the methods of Figs. 3 to 6 are effected by instructions in the software that are carried out by the computer. The software may be divided into two separate parts; one part for carrying out

the methods, and another part to manage the user interface between the latter and the user. The software may be stored in a computer readable medium, including the storage devices described below, for example. The software is loaded into the computer from the computer readable medium, and then executed by the computer. A computer readable
5 medium having such software or computer program recorded on it is a computer program product. The use of the computer program product in the computer preferably effects an advantageous apparatus for formatted document generation in accordance with the embodiments of the invention.

The computer system 700 comprises a computer module 701, input devices such
10 as a keyboard 702 and mouse 703, output devices including a printer 715 and a display device 714. The display 714 is used to reproduce the GUI and images depicted in Figs. 1, 2A-2C, 3 and 5, whilst the printer 715 may be used to print the printable documents 300 and 500 of Figs. 3 and 5 respectively. A Modulator-Demodulator (Modem) transceiver device 716 is used by the computer module 701 for communicating
15 to and from a communications network 720, for example connectable via a telephone line 721 or other functional medium. The modem 716 can be used to obtain access to the Internet, and other network systems, such as a Local Area Network (LAN) or a Wide Area Network (WAN).

The computer module 701 typically includes at least one processor unit 705, a
20 memory unit 706, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a video interface 707, and an I/O interface 713 for the keyboard 702 and mouse 703 and optionally a joystick (not illustrated), and an interface 708 for the modem 716. A storage device 709 is provided and typically includes a hard disk drive 710 and a floppy disk
25 drive 711. A magnetic tape drive (not illustrated) may also be used. A CD-ROM drive 712 is typically provided as a non-volatile source of data. The components 705 to 713 of the computer module 701, typically communicate via an interconnected bus 704 and in a manner which results in a conventional mode of operation of the computer system 700 known to those in the relevant art. Examples of computers on which the

embodiments can be practised include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom.

Typically, the application program of the described embodiments is resident on the hard disk drive 710 and read and controlled in its execution by the processor 705. Intermediate storage of the program and any data fetched from the network 720 may be accomplished using the semiconductor memory 706, possibly in concert with the hard disk drive 710. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk and read via the corresponding drive 712 or 711, or alternatively may be read by the user from the network 720 via the modem device 716. Still further, the software can also be loaded into the computer system 700 from other computer readable medium including magnetic tape, a ROM or integrated circuit, a magneto-optical disk, a radio or infra-red transmission channel between the computer module 701 and another device, a computer readable card such as a PCMCIA card, and the Internet and Intranets including e-mail transmissions and information recorded on websites and the like. The foregoing is merely exemplary of relevant computer readable mediums. Other computer readable mediums may be practiced without departing from the scope and spirit of the invention.

The methods of the described embodiments may alternatively be implemented in dedicated hardware such as one or more integrated circuits performing one or more functions or sub-functions of the formatted document generator. Such dedicated hardware may include graphic processors, digital signal processors, or one or more microprocessors and associated memories.

Industrial Applicability

Embodiments of the present invention are applicable to network data accessing and retrieval systems and the described embodiments are intended to complement existing browsing and searching tools, particularly in Internet and World Wide Web applications. It is also noted that the embodiment of Figs. 5 and 6 is not limited to for use with the embodiment of Figs. 3 and 4, but finds general application for use with the Hypertext Document Collating Tool, and also in like arrangements. Further, whilst the embodiment

of Figs. 3 and 4 describes an arrangement that is adjunct to the search engine application, an alternative embodiment may be integrated into the search engine application so as to automatically supplement or replace, the provision of search results as depicted in Fig. 1. For example, the search engine may be configured to generate a file which incorporates
5 the search string, the URL's returned in the search result, and user preferences for the formatted document generation. Such preferences may include the number of URL's to be examined for any one search result (eg. the first ten), the amount of information to be extracted from any referenced URL (eg. the top of page, the number of hits on the searched string, the size of the extracted string containing portion), and a maximum size
10 (eg. in terms of printable pages) of the printable document for any one searched string, to name but a few. The provision of such a file, and appropriate configuration of the formatted document generator application to interpret the file, can facilitate the batch processing of search strings and corresponding document generation without user interaction.

15 Further, with respect to the embodiments of Figs. 5 and 6, rather than being arranged within columns, the advertisements 522 and/or 524 may be arranged to extend across the width of the page, for example like a "banner" style advertisement.

Also, whilst the embodiment of Figs. 3 and 4 has been described with reference to Internet and World Wide Web searching, the inventive concept is not limited thereto,
20 but applies generally to all computer networks. For example, a Local Area Network within an office environment may incorporate many thousands of word processing documents distributed amongst many computer devices. The described embodiments may thus be used to perform keyword searching on such documents in order to identify specific classes of documents without the user having to open each individual document.

25 The foregoing describes only some embodiments of the present invention, and modifications and/or changes can be made thereto without departing from the scope and spirit of the invention, the embodiments being illustrative and not restrictive.